Prince Declaration
Exhibit 6
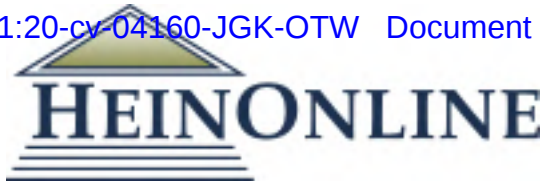
**Exhibit
0003**

Can Private Copyright Protection Be Effective? Evidence from Book Publishing

Imke Reimers   *Northeastern University*

Abstract

Digitization has impacted publishing, news, and entertainment industries in re-
cent years by lowering the cost of access. With the option to download creative
works legally, however, come the possibility of doing so illegally and the issue
of how to protect copyrighted works effectively. Public (legislative) and for-
mal (legal) efforts to prevent copyright infringements have been controversial
or inefficient. The book industry showcases an alternative approach in which
private companies use relatively inexpensive network surveillance to protect in-
dividual titles. I estimate the effectiveness of such protection on legal sales of
books that become protected using a difference-in-differences approach. I find
a protection-related increase in sales of electronic books—the closest substitute
for online piracy—of more than 14 percent, with effectiveness depending on
popularity, genre, and search frequency. Most of the increase is due to preven-
tion of casual infringements rather than professional piracy.

1. Introduction

Digitization has significantly decreased the costs of creating and distributing cul-
tural goods in media industries, and it has started to do the same in other indus-
tries as well (for example, using three-dimensional [3-D] printing). The collapse
of traditional costs has increased the quantity and variety of products available to
consumers. The wider variety is expected to increase the level of competition and
lower prices further. Consequently, it is affecting firms' and artists' profitability.
By themselves, these changes shift surplus from producers to consumers.

Digitization has had another effect though. As the legal distribution of cre-
ative works has become cheaper and easier, so has the illegal distribution of these

works. Recent academic studies show that illegal distribution displaces legal sales in media industries, even though modern technology has made monitoring and regulation of illegal activity more feasible. Thus, the question of what is the best intellectual property strategy becomes extremely important for each artist, firm, and distributor and in each industry. To what extent can works be effectively protected from piracy, what should this protection look like, and who should enforce it?

Many well-known efforts to protect copyright have been public, or legislative, in nature, with mixed success and reception. Other efforts have been legal in nature, often attempting to protect the entire industry at once—an effort that can be very costly. A less formal type of protection, private copyright protection that targets individual works, has received less attention but might have clear advantages. Most important, the ability to target a subset of works could make the effort to prevent infringement more efficient.

The book publishing industry provides a setting that allows me to analyze such an effort to protect a work's copyright. In the book publishing industry, targeted copyright protection efforts are led by private companies that are hired by publishers to search the Internet for infringing content specific to protected titles. Protection from piracy is title and publisher specific rather than industry wide. It is initiated by publishers and authors rather than by public officials, and the firms attempt to remove infringing content without the involvement of courts, which further lowers the costs of protection.

I estimate the effect of such a private antipiracy effort on legal book sales. I use a novel data set consisting of monthly electronic book (e-book) sales of titles that are offered exclusively in electronic format by one publisher (RosettaBooks) and those titles' weekly physical sales from January 2010 to December 2013. The data set includes the intensity level and success rate of protection from piracy (hereafter, piracy protection) over the same time period through Digimarc, a large piracy protection company. I follow sales of each title in the publisher's catalog before and after its title-specific piracy protection begins in a difference-in-differences setting. I check whether (and by how much) sales rise or fall when protection for a title begins relative to the sales of other titles that become protected at another time.

In this analysis, I take advantage of the facts that initiating protection of a work is a relatively labor-intensive process that was implemented in shifts over several months by Digimarc and that RosettaBooks seemed unaware that the order in which titles were submitted might affect its profits. I further use data from Google Trends to provide evidence that the timing of protection is exogenous and consistent. Here I take advantage of the fact that the Internet search volume likely reflects consumers' interest in a work, whereas it should be uncorrelated with the timing and level of piracy protection, at least before protection begins.

A large and growing literature addresses the effect of file sharing in media industries (see Smith and Telang [2012] for a detailed description; see Peitz and Waelbroeck [2006a] for a review of the theoretical literature). While some early

work indicates that there is no significant effect (Oberholzer-Gee and Strumpf 2007), most recent work finds that legal sales are significantly displaced by pirated versions (Zentner 2006; Rob and Waldfogel 2006; Liebowitz 2008; Waldfogel 2010; Danaher, Smith, and Telang 2015). In the music industry, the negative effect of file sharing on legal sales can in part be attributed to the arrival of Napster in 1999, a file-sharing website that revolutionized the industry. The first legal option to download music did not appear until almost 4 years later, when Apple sold its first songs through the iTunes Store in May 2003.

More recently, research on file sharing has turned to determining whether antipiracy efforts can be effective, taking as a given that piracy likely harms sales in an industry. Most efforts to protect creative works have been legislative, and many have been controversial. Consider, for example, the Internet blackout to campaign against the Stop Online Piracy Act (SOPA; H.R. 3261, 112th Cong. [2011]) and the Protect Intellectual Property Act (PIPA; S. 968, 112th Cong. [2011]) in January 2012.[1] Similarly, the French Parliament passed the Creation and Internet Law, an antipiracy law more commonly known as HADOPI, in 2009 (Loi 2009-669 du 12 juin 2009 favorisant la diffusion et la protection de la création sur internet [Journal Officiel de la République Française [J.O.] [Official Gazette of France], June 13, 2009, p. 9666]). This bill introduced a three-strikes policy in which consumers of illegal music were blocked from the Internet after two warning notices. While Danaher et al. (2014) find that awareness of this law increased legal music sales on iTunes, the bill faced controversy as France's Constitutional Council declared access to the Internet a basic human right, and the bill was finally revoked in July 2013.

Most legal efforts to prevent piracy tend to be broad in scope as well. In the movie industry, the abrupt shutdown of the cyberlocker Megaupload in January 2012 had a significant positive effect on box office revenues and digital movie sales of popular works, but the shutdown did not have a positive effect on box office revenues of less well-known works (Peukert, Claussen, and Kretschmer 2013; Danaher and Smith 2014). In music, the Recording Industry Association of America made legal threats against file sharers, which decreased the level of file sharing, although a substantial number of illegal files remain available (Bhattacharjee et al. 2006). Another effort by the major music labels to control secondary distribution of content through digital rights management (DRM) does seem to have a positive effect on sales (Zhang 2014).[2]

Legal efforts to prevent copyright infringement face less controversy than legislative efforts, but it is not clear if the gains from such protection outweigh the effort that is exerted to protect a work, especially when protection is broad. I address these questions by examining the effectiveness of a less formal form of pi-

[1] The acts aimed to facilitate the combating of online copyright infringement by (among other measures) barring advertising networks from conducting business with infringing sites, barring Web search engines from linking to those sites, and requiring Internet service providers to block access to them.

[2] Digital rights management restricts consumers' ability to copy (and thus redistribute) the content they have purchased.

racy protection—one in which private firms offer surveillance of piracy activity for specific titles—by comparing the resulting increases in revenue with the costs of protection. My work thus lends support to Luo and Mortimer (2015), who find that private enforcement in the form of ex post infringement settlement requests for stock photography can indeed be successful. I show that these attempts can lead to increases in revenue that outweigh the cost of protection, especially when legal action can be avoided.

In particular, I find that the effect of piracy protection on legal sales of books depends on the popularity of the title, the type of work, and the format of the edition. While physical formats are not affected by piracy protection, closer substitutes for online piracy such as legally distributed e-books see a mean differential protection-related increase in sales of at least 14 percent. My results suggest that the effect is driven more by the delisting of links from search engines than by a decrease in piracy sites offering the titles. This indicates that casual offenders are more easily deterred than serious pirates. Moreover, this type of piracy protection increases e-book revenue by up to 10 times the cost of protection. I find that private and targeted piracy protection can be effective, and if the right set of works is protected with an appropriate level of intensity, this protection can even be cost-efficient.

The remainder of the paper proceeds as follows. Section 2 provides background on the publishing industry and the role of piracy. I describe the data in detail in Section 3, and I describe my empirical strategy and provide evidence that the effect is identified in Section 4. I show results for the effectiveness and efficiency of protection in Section 5, and I examine the mechanism by which protection works in Section 6. I conclude with policy implications in Section 7.

2. Industry Background and Piracy Protection

The recorded-music industry faced challenges from digitization when the Napster file-sharing service was introduced in 1999. Since then, digitization has posed challenges to all media industries, and it will likely pose challenges in nonmedia industries as well, as digitization (for example, 3-D printing) becomes more widely used. Some of these challenges can be attributed to increased competition through free (pirated) versions.

The effects of file sharing and the effort to stop it are not yet well studied for book publishing, as digitization in this industry has lagged behind other media industries.[3] While e-books can be read on computers and have therefore been available for over a decade, they are most useful to consumers when read on small hand-held devices, such as e-book readers or tablet computers. Sony released e-book readers as early as 2004, but the first widely adopted e-reading de-

---

[3] Hardy, Krawczyk, and Tyrowicz (2014) report on a field experiment in which they limited piracy for a set of more recent book titles in Poland. While they find that piracy can in fact be limited, they do not find an effect of such action on legal sales. My paper looks at a different set of titles and a slightly more comprehensive protection strategy. While their strategy consists only of takedown notices, my strategy includes requests to delist links from search engines.

vice was Amazon's Kindle, which was introduced in November 2007, when there was no widespread option to download content illegally. The e-reader and e-book markets have grown quickly since then. The share of US adults owning an e-book reader grew from 2 percent in April 2009 to 24 percent in September 2013, and the share of adults owning either an e-reader or a tablet (or some viable method for consuming e-books) reached 50 percent in January 2014 (see Sterling 2014). Electronic books have similarly become increasingly popular. While e-books held a negligible market share among fiction books before 2007, about half of the weekly *USA Today* top 150 best-selling books had been sold primarily as e-books by late 2012 (see Waldfogel and Reimers 2015).

Online piracy has been relatively concentrated in other content industries. Although many book titles are available on well-known piracy sites (for example, peer-to-peer networks such as Piratebay), the book industry does not face one large file-sharing service, as was the case with Napster in the music industry.[4] Book files are smaller than music and movie files, which makes it convenient to store e-books as intact files. Thus, much of the infringing content is located on websites rather than on peer-to-peer networks, which makes tracking and removing illegal content more feasible.

The small file size of books may also lead to fewer dedicated pirates. Most book content (65 percent) is either in HTML or pdf format—formats that are easy to upload and download but can detract from the reading experience. Only 20 percent of the infringing content in my data set is clearly recognizable as being in an e-book format, perhaps because these formats have additional protection against copying (for example, DRM).[5]

In the book publishing industry, copyright protection involves two major steps: requesting that search engines delist the link to the infringing content and demanding the site's host remove the pirated content directly. The two actions target different types of pirates. Consumers who are deterred by changes to search results are likely casual pirates, while more serious pirates are less likely to rely on search engines to find content. Sivan, Smith, and Telang (2014) show for the movie industry that the presence of links to pirated material in search results strongly affects the piracy behavior of consumers.

This paper follows titles that become protected by Digimarc, one of the industry's largest antipiracy services.[6] Its piracy protection strategy includes an automated crawling process to find suspected pirated content. Sites with confirmed pirated content are subject to two types of treatment: the company sends requests to the search engines Google and Yahoo! to delist the links, and it sends takedown notices to the provider of the illegal content, essentially until the content is

---

[4] Peer-to-peer networks are created when two or more computers are connected and share resources without going through a separate server computer.

[5] It is not obvious why pirates offer infringing content for books, as offering this content is not typically monetized. Many suppliers of pirated content list entertainment, a sense of community, and popularity as the main motivators for offering such content.

[6] The service was introduced by Attributor, which was acquired by Digimarc in December 2012. It is now known as Digimarc Guardian.

removed. These two actions affect different types of pirates. Delisting from search engines prevents casual consumers of illegal content—those who rely on search engines to find illegal content—from reading pirated versions, while removing infringing material has a better chance of diverting more experienced pirates—those who know where to find pirated content anyway. Protection continues as long as the title is under contract.

Digimarc has added most of the major US publishing houses as clients over the past few years. The publishing companies include HarperCollins and Macmillan (both signed in 2010, although Macmillan now uses another service), Simon & Schuster (2011), and Random House (2013, now Penguin Random House). In addition, it added several smaller publishers. RosettaBooks, which signed with the company in June 2011, is one of those publishing houses.

RosettaBooks has secured exclusive rights to publish electronic versions of over 600 titles.[7] Its list consists largely of backlist titles, ranging from well-known classics (including, among others, works by Kurt Vonnegut, Arthur C. Clarke's *Space Odyssey* books, and Stephen Covey's *The Seven Habits of Highly Effective People*) to works that are less well known today. While most of its titles were originally published more than a decade ago, RosettaBooks also carries some original titles that are available only in electronic format.

Protection for RosettaBooks' full catalog was due to begin in June 2011. Although the entire catalog came under protection on the same date, the implementation occurred in shifts, and the timing was not coordinated with the publisher. When asked why titles became protected in shifts, an employee of Digimarc relayed that the timing with which titles came under protection was related to capacity constraints rather than strategic decisions.[8] In addition, according to this account, the publisher was not aware of any capacity constraints at the time.[9]

3. Data

I follow the demand for and piracy protection of a set of book titles over 4 years, from 2010 to 2013. The underlying data set consists of 653 titles whose electronic versions are exclusively published through RosettaBooks.[10] Most of the works were originally published several years ago, going back as far as the first

---

[7] While electronic editions are published only by RosettaBooks, physical editions are released by other publishers.

[8] The employee stated: "I am afraid we do not have concrete reasons why titles were protected in shifts. . . . It took some doing before we had a regular ONIX feed set up and it appears that the global catalog was added in shifts, where large blocks were ingested in batches rather than the entire catalog being added in one shot. I think we actually got a flat file or two before the full feed, and then deltas following the initial large feed with anything that wasn't on the first list. Why it was broken down that way isn't something we have much explanation for, apart from saying this is how it was supplied to us, and it took a while before it was complete" (Blair Elefant, Digimarc senior relationship manager, e-mail correspondence with the author, July 28, 2015).

[9] I am especially grateful for the recommendations of an anonymous referee who suggested a number of econometric tests to establish exogeneity of the timing in support of this e-mail message. These tests and the affirmative results are described in Sections 4 and 5.

[10] For the full catalog, see RosettaBooks, Ebooks A–Z (http://www.rosettabooks.com/ebooks-a-z/).

Private Copyright Protection   417

Table 1
Effect of Takedown Notices

	N	Mean	SD	Min	Max
Sites per title	251	87.899	183.025	1	1,984
Notices per site	20,041	1.596	1.259	0	29
Success	20,041	.835	.371	0	1
Success after first notice	20,041	.745	.436	0	1
Time until success (days)[a]	16,739	17.834	56.433	0	857

[a] For sites with infringing content that was successfully removed.

half of the 20 century, although all titles are still protected by copyright. There also are a few titles that were published in the past 15 years. My analysis includes the subset of 251 titles for which pirated content was found between 2011 and 2013.

On the demand side, I use the Nielsen BookScan database to observe weekly sales data for physical book formats: hardcover, trade paperback, mass-market paperback, and audio versions. While these sales data date back to 2002, I focus on the time period from January 2010 to December 2014. These data include all editions of the title that are listed by Nielsen. I also observe monthly e-book unit sales at the title level directly through RosettaBooks (the exclusive publisher of the titles' e-books) from July 2011 to December 2013. These editions are sold through Amazon, Apple's iBooks Store, and Barnes and Noble. The publisher carries a few popular titles that sell far more than 100,000 copies annually and some that sell fewer than 1,000. On average, editions sell close to 3,000 copies a year, although this varies by format, with trade paperbacks and e-books being the most popular formats.

On the piracy dimension, I obtained detailed information from Digimarc. I observe when the title came under contract to be protected, the date that each illegal link to each title was found, when takedown notices were sent to websites offering specific titles, and when the pirated content was taken down. The titles in my data set have 24 different protection start dates between August 2011 and November 2013. Digimarc searches for titles at three different frequency levels, depending on the amount of piracy that is found. The search frequency of a title can vary over time, although it does not do so often.

Table 1 shows summary statistics for all titles that were protected during the observed time period. The antipiracy service appears to have some success in reducing file sharing, and this success comes reasonably quickly. Over 50 percent (10,417) of all infringing content is taken down within 1 week of being found. Yet it is possible that not all websites are found and that taking down content on one website will simply lead to another one showing up to offer the same title (on the emergence of new piracy sites when content on one is removed, see Aguiar, Claussen, and Peukert [2015] and Danaher, Smith, and Telang [2015]). The fact that Digimarc finds close to 88 sites per title on average, and that pirated content on over 83 percent of those is successfully removed, indicates that the company
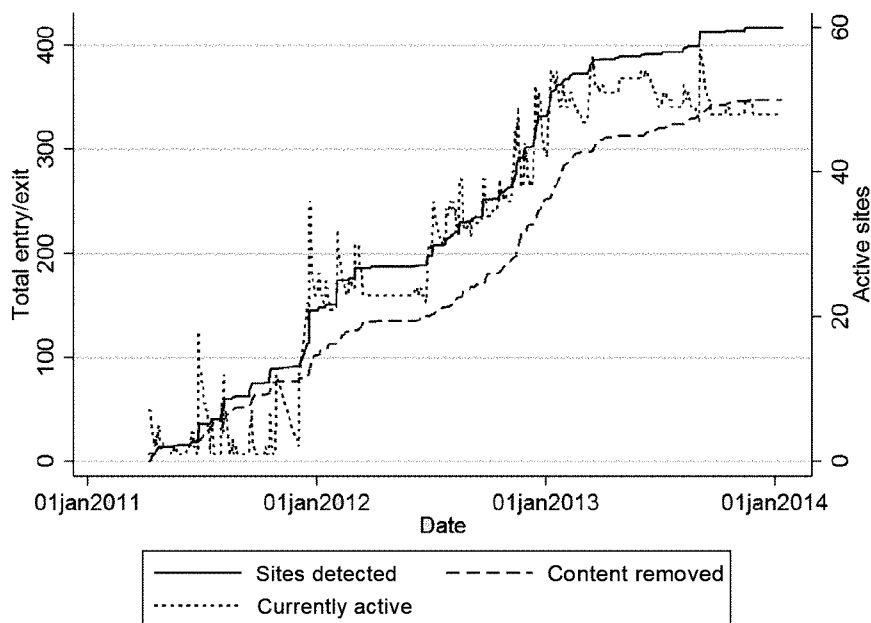
Figure 1. Piracy of *I Am Legend*

affects the online book piracy landscape, at least for the titles that RosettaBooks offers. The relatively high success rate may be due to the nature of piracy in book publishing. Because e-book files are small, there are more small sites with infringing material, and the operators of such sites may be more easily intimidated by takedown notices.[11] Moreover, since offering pirated content is not usually monetized, pirates have no direct incentives to continue offering the content.

The number of detected piracy sites with infringing content that has not been removed increases on average over time because not all takedown notices are successful. Figure 1 illustrates the statistics for Richard Matheson's *I Am Legend*.[12] The number of active sites decreases at times. For example, while there is a spike in piracy in late December 2011 with up to 36 infringing sites (December 21), some of the copyrighted content on these has been successfully removed by January 2012, with only 23 active sites by January 1. In early 2014, though, 50 infringing sites can be found. This variation allows me to use the net change in the number of sites as a measure of the success of takedown notices.

[11] Overall, 69.8 percent of infringing content in my data set is found on cyberlockers (Internet hosting services designed to host users' files), most of which is in HTML and pdf formats, while only 18 percent is found on peer-to-peer networks. Peer-to-peer networks are notoriously difficult to contain, with a success rate of only 65 percent, compared with 87 percent for cyberlockers.

[12] The facts that I do not observe the number of sites with infringing content before protection and that the number of active sites detected increases over time make it difficult to use protection as an instrument for the level of piracy. I instead estimate the effect of protection itself.

The Google Transparency Project and the website Chilling Effects (now Lumen) track which links have been subject to delisting requests. Most of the links are in fact removed from the search engine. Between May 2013 and September 2014, only 358 of Digimarc's 10,198 requests (3.5 percent) regarding RosettaBooks were not honored. However, there is no information about which links remained active and, more important, when the link was delisted, although the delisting typically takes place within hours.

I use two more sources to collect additional information about each title's interest level (popularity) and quality. To measure each title's popularity at any given time, independent of its piracy level, I collected the Google search history for all titles and authors from January 2011 to December 2013 through Google Trends. While Google Trends does not provide absolute counts of search queries, it allows me to compare the relative interest levels across time and across search queries.[13] I collected information about the quality of titles using the crowd-sourced online book database Goodreads. The mean (median) number of reviews for the titles in my analysis is 4,855 (483) for e-books and 20,821 (1,500) for physical books, and the mean (median) star rating (on a scale from 1 to 5) is 3.75 (3.75) for e-books and 3.85 (3.86) for physical books. I use this information when disentangling the effect of piracy on different types of titles. Finally, I use information about site traffic from Alexa Internet, Inc., to approximate the actual level of piracy in robustness checks.

4. Empirical Strategy and Identification

In an ideal experiment that determines the effect of piracy protection, the econometrician assigns titles randomly across groups and protects one set of titles. I observe a quasi experiment in which the timing of protection adoption across titles is uncorrelated with the returns to protection. Although protection for all of RosettaBooks' titles was contracted for at the same time, the order in which the protection was applied was determined by when each title was submitted, and the order of submission was independent of a title's characteristics.[14] The marginal effect of online piracy protection on book sales is then the change in sales of the treated titles as compared with the change in sales of the group of titles that were not yet treated. The empirical strategy takes advantage of the change in protection in a difference-in-differences analysis. The log of the (observed) unit sales q_{it} of title i in time period t is a function of the work's observable characteristics and its protection status in that time period.

The effect of piracy protection is identified because I observe sales of titles when they are protected and when they are not protected. Since implementation of piracy protection occurred at different times for different titles, it is unlikely that any changes in sales when a title moved into piracy protection are due to

---

[13] I use the variation on those dimensions to support my claims of exogeneity of protection in Section 5.4.

[14] I provide evidence of this independence in Section 5.4.

an exogenous time-dependent shock that affected the overall demand for books. Moreover, a change in the level of piracy for one title is unlikely to directly affect legal demand for another title (although it might affect the level of piracy for other titles). I estimate the effect of piracy protection as

$$\log(q_{it}^k) = X_{it}'\beta + \alpha \times \text{Protection}_{it} + \delta_i^k + \mu_t^k + \varepsilon_{it}, \tag{1}$$

where q_{it}^k denotes the unit sales of title i in week t in format k, with $k \in$ {hard-cover, trade paperback, mass-market paperback, audiobook, and e-book}, and X_{it} includes time-varying characteristics of titles including recent editions, age, average sales prices, and the search volume on Google (the readers' interest in the title).[15] Finally, Protection_{it} is a dummy variable that equals one if the title is under piracy protection in week t. The effect of piracy protection on legal book sales is given by α, where piracy protection causes a change in legal sales of $e^\alpha - 1$ percent.[16] The log specification assumes that piracy changes sales by a common percentage rather than by a common absolute amount. This controls for differences in absolute sales levels across titles.

I also make use of the fact that I observe unit sales of over 120 titles (depending on format) over an extended time period (170 weeks for physical books and 30 months for e-books) by controlling for title i and format k (δ_i^k) and by adding time fixed effects (μ_t^k), which are allowed to vary across formats (e-books have become more popular over time). The title fixed effects pick up the overall popularity of the title, the genre, and the author. Time fixed effects control for changes over time that affect all titles similarly, such as changes in the economic environment or the release of a new e-reader or tablet.

In additional specifications, I use the overall protection and a measure for the effectiveness of takedown notices as my independent variables of interest to identify the leading mechanism behind the success of piracy protection. I also estimate the role of intensity of protection: the effect of different levels of search for infringing content. I then separate the effect of piracy protection on sales of well-known works and less well-known works and sales of fiction and nonfiction titles to determine what types of works are more likely to benefit from protection.

5. Results

I follow unit sales for different formats of the titles from January 2010 to January 2014 (physical sales) and from July 2011 to December 2013 (e-book sales). I estimate several versions of equation (1) for different formats and sets of titles.

---

[15] I control for readers' interest instead of using an instrumental variables approach in which I estimate a title's protection status as a function of readers' interest in the first stage because readers' interest is not a good predictor of a title's protection status.

[16] As I observe some periods with no physical sales for some titles, I use $\log(q_{it}^k + .00001)$ as the dependent variable for physical formats. In that case, the marginal effect is interpreted with a bias. The bias understates the true effect but goes to 0 quickly (the bias is $\Delta q/q(q + .0001)$). The estimation of e-book sales does not require this adjustment, as I do not observe time periods with no sales.

Private Copyright Protection   421

Table 2
Baseline Results: Titles Moving into Piracy Protection

	Hardcover	Paperback		Audio	E-Book
		Trade	Mass Market		
Protection	−.0649	−.167	.103	.342	.144+
	(.338)	(.192)	(.136)	(.415)	(.0768)
Time fixed effects	Weekly	Weekly	Weekly	Weekly	Monthly
Observations	9,350	13,506	8,714	6,630	3,104
R^2	.024	.043	.039	.041	.081
Titles	60	88	59	43	126

Note. Robust standard errors are clustered by title. Results for the dependent variables are in logs of
unit sales. The percentage change in unit sales as a result of protection is calculated as $e^{\beta \text{Protection}} - 1$. All
regressions include title fixed effects.
  + $p < .1$.

5.1. Baseline: Piracy Protection

Table 2 shows the baseline estimation results for all titles that moved into pro-
tection at some point during the observed time period, with at least 2 months on
and off protection. The independent variable of interest is an indicator variable
that equals one if the title is under contract with Digimarc. I treat a title as pro-
tected in all time periods after the first piracy site is found, regardless of the num-
ber of sites that are found and/or from which content is taken down in any given
time period. The dependent variable is the log of unit sales of format k of title i in
time period t. Note that the number of titles varies across formats because some
titles are not available in all formats.

Table 2 shows that the effect of piracy protection on legal book sales depends
on the edition's format. Electronic books can be regarded as the closest substi-
tutes for pirated versions. Piracy protection has a large and statistically significant
positive effect (at 10 percent) on sales of e-books. Moving a title into piracy pro-
tection increases e-book sales by 15.5 percent ($=e^{.144} - 1$).[17] Physical editions of
a title, on the other hand, are not necessarily close substitutes for free electronic
versions of the same title. Consequently, the effect of piracy protection is insig-
nificant at the 10 percent level for the physical formats (including audiobooks)
and negative (but highly insignificant) for hardcover books and trade paperbacks.
More generally, the effect on physical formats is estimated very imprecisely,
which leads me to focus much of my analysis on e-books.

---

  [17] The remainder of the paper reports the extrapolated percentage change in sales, while the tables
report the estimated coefficients. Note that I cluster standard errors by title to account for common
group effects and to reduce the potential for overstating significance due to serial correlation (see
Bertrand, Duflo, and Mullainathan 2004). The number of titles is large, so any nonnormality of the
errors is not a big concern (see Donald and Lang 2007). Note also that this positive effect of protec-
tion is robust to different definitions of protection. See Appendix Section A1.

5.2. Additional Controls

Of course, Table 2 does not take into account several factors that may contribute to an increase in e-book sales around the time that a title becomes protected. It could be that a move into protection coincides with the publication of a new edition of that title, with title- and age-related fluctuations in demand, or with price promotions.

A new edition of a book can direct attention to the title and spur demand. An increase in legal sales is expected even if we also expect more piracy. But if a publisher releases a new hardcover edition, competition for the other formats increases, which has a negative effect on sales of those formats. To determine if legal sales (and hence the inferred effect of piracy protection) are driven by recently published editions or by the title's age, I add three indicator variables that equal one if a new electronic, print, or audio edition was published in the 2 months prior to the month in question and a quadratic function of title's age (in months) to equation (1). I further account for title-specific fluctuations in demand by controlling for the title's and author's Google Trends search volume in the estimation and for the total number of active piracy sites detected.

Finally, the publisher can adjust a book's price. Oestreicher-Singer and Sundararajan (2010) show that DRM has an economically significant effect on pricing decisions for digital content, and it is possible that piracy protection does as well. Since preventing piracy effectively limits the level of competition in the market, theory predicts a resulting increase in prices. If this is the case, then the positive effect of piracy protection on unit sales would be countered by a negative effect of higher prices on unit sales, and the above results would underestimate the true effect of piracy. If, on the other hand, prices decrease when a title is protected, then the results reported above will be overestimated: my estimates would reflect a combination of piracy protection and lower prices.[18] An unreported difference-in-differences analysis with e-book prices as the dependent variable indicates that e-book prices decreased by 5.3 percent as a result of a move into piracy protection. This effect is significant at the 10 percent level. A negative correlation between prices and protection would indicate that the results above overestimate the effect of piracy protection. I therefore control for prices in the analysis as well.

Table 3 shows that a recently published electronic edition increases e-book demand for a title, regardless of the specification. The effect of piracy protection on e-book sales remains similar in size and significance as I add more control variables. Even when I control for the average sales price (column 3), the effect of pi-

[18] A decrease in prices as the level of competition decreases is counterintuitive at first, but it is possible. When a zero-price file-sharing option is available, the publisher of the higher-quality, non-zero-price option may target high-valuation consumers exclusively. Without the low-quality competitor, the publisher can target all types of consumers but may have to lower its price to reach the low-valuation consumers. See Petrin (2002) for a structural analysis of this effect in the car industry. Alternatively, piracy protection increases the publisher's awareness of the zero-price competitor in the first place.

Private Copyright Protection   423

Table 3
Effect of Piracy Protection on E-Book Sales

	E-Book Sales			E-Book Revenue
	(1)	(2)	(3)	(4)
Protection	.153+	.177*	.147*	.157*
	(.0775)	(.0830)	(.0613)	(.0638)
Recent e-book	.244*	.273**	.388**	.351**
	(.104)	(.103)	(.0785)	(.0811)
Recent print	−.366	−.371	.160	−.0130
	(.320)	(.335)	(.175)	(.199)
Recent audio	.247	.256	.270+	.265
	(.249)	(.238)	(.157)	(.174)
Title search volume		.0154	.00911	.0112
		(.00934)	(.00714)	(.00776)
No title interest		−.139	−.109	−.119
		(.359)	(.258)	(.287)
Author search volume		.0000571	−.000661	−.000427
		(.000812)	(.000861)	(.000818)
No author interest		−.203**	−.394**	−.332**
		(.0638)	(.105)	(.0863)
Active sites		−.0198+	−.0108	−.0138
		(.0113)	(.00876)	(.00851)
Title age (months)		.0229*	−.00611	.00332
		(.00869)	(.00646)	(.00637)
Title age^2		−.00000857	.0000263**	.0000150*
		(.00000929)	(.00000719)	(.00000746)
Log price			−1.483**	
			(.155)	
Observations	3,104	3,049	3,049	3,049
R^2	.084	.091	.480	.127
Titles	126	121	121	121

Note. Robust standard errors are clustered by title. Results for the dependent variables are in log e-book sales or log e-book revenue. Recent editions equal one if an edition was published in the previous 2 months. Title and author search volume data are normalized such that Arthur C. Clarke's *2001: A Space Odyssey* has a search volume of 100 in January 2011. Variables for no title and no author interest equal one if there was not enough search volume for Google Trends to record a number. All regressions include title and monthly time fixed effects.

+ $p < .1$.
* $p < .05$.
** $p < .01$.

racy protection remains similarly strong and significant at the 5 percent level and leads to a 15.8 percent increase in unit sales. Column 3 also shows that consumers are relatively price elastic with a price elasticity of −1.48. I also estimate the impact of protection on e-book revenue and find that e-book revenue increases by 17 percent when a title becomes protected.

The remaining coefficients are reasonable as well. One would expect e-book sales to increase when a new e-book edition is published, while a new print edition (a competing format) could move demand away from the electronic edition.

Audio editions, on the other hand, may serve as complements to e-books.[19] Moreover, when a title's Google search volume is higher, it tends to sell more units. The negative coefficients on the number of active piracy sites confirm that these sites provide substitutes for legal e-book sales. After year-month fixed effects are added, the fact that the titles in the data set become older over time does not seem to affect sales much. The positive coefficient on the square of the title age suggests that demand for more recent titles is more likely to decrease in the time period of the analysis, while demand for older titles leveled off even more by 2011.

5.3. Costs, Benefits, and Efficiency

Private copyright protection can indeed be effective, at least for the e-books in my data set. Furthermore, piracy protection can be efficient in the sense that the benefits to the publisher, authors, and the piracy protection service outweigh their costs. In 2010, RosettaBooks' total revenue was over $1.5 million across more than 300 books (see Deahl 2011). Since then, the number has likely increased significantly, as e-books have become more popular and RosettaBooks has substantially increased its catalog. Taking the annual revenue of $1.5 million as a given, I find that RosettaBooks' revenue would increase by $255,000 if all works benefited equally from piracy protection. Of course, pirated content was found for only 228 titles (of 673), so RosettaBooks' total revenue likely increased by less than that. Since the reported revenue of $1.5 million corresponds to a catalog of approximately 300 titles, one might assume that the incremental revenue from the 228 titles for which piracy was found is $193,800 ($=255,000 \times 228/300$).[20]

This benefit needs to be compared with the costs of protection. To the publisher, this cost is the fee it pays to the antipiracy company—a simple transfer of wealth. For the antipiracy company, the marginal cost of protecting the publisher consists of several dimensions in addition to the fixed costs of running the company. The publisher-specific marginal costs of protection include the development of scraping scripts, manual discovery of infringing content (by emulating the behavior of pirates), approval of the links that have been identified as infringing (for example, to negotiate fair-use situations), and enforcement (sending notices, monitoring, legal escalation, and so on). I can make inferences about these publisher-specific costs, as RosettaBooks' titles account for .26 percent of all links that Digimarc found in the previous year, which amounts to 823 hours of incremental operational work to protect the publisher.[21]

The revenue to the publisher and authors (as royalties) increases by approximately $194,000. At an hourly wage at the US average of around $25, the incremental cost to the antipiracy company of protecting the publisher's titles

[19] For example, one might listen to an audiobook in a car and continue to read the e-book version at home.

[20] This is a conservative estimate, as the pirated titles are likely more popular than those that are not pirated.

[21] Of course, this measure does not include Digimarc's overhead and fixed costs of operating. The true amount of time needed to protect the RosettaBooks catalog would be larger.

amounts to $20,575—not much more than 10 percent of the increase in revenue. At the minimum wage, this cost would of course be even lower. These calculations suggest that the benefits to the publisher and authors outweigh the marginal costs to the antipiracy company by a large margin. Whether these benefits also outweigh the fixed (overhead) costs of running the antipiracy company and the potential losses in consumer surplus due to the removal of the zero-price pirated option is beyond the scope of the available data, although it is reasonable to assume that the overhead costs associated with a surveillance operation of this sort are small, especially when they are distributed over several publisher-clients.

5.4. Evidence Supporting the Exogeneity of Protection

The identification strategy assumes that the timing of piracy protection is uncorrelated with factors that determine the outcome of interest—in this case, the log of unit sales. Of course, a publisher may be more interested in protecting titles that have recently gained popularity than in protecting those that do not attract much interest from readers. For example, RosettaBooks would have liked to increase piracy protection for Richard Matheson's 1954 novel *I Am Legend* when it was adapted as a blockbuster movie in 2007. The change in legal book sales would then be the result of a combination of two effects: an increase in demand due to the movie's promotional effect on the novel and the change in demand from a change in piracy protection.

I deal with this problem in several ways. First, in addition to controlling for a large list of demand shifters, I use a list of older titles—most were published before 2000—so it is unlikely that a change in demand is caused by title-specific market shocks. Second, conversations with representatives of RosettaBooks and Digimarc suggest that the timing of protection was not determined by the publisher. Third, I use Google Trends data to provide two types of evidence that the timing of protection is not correlated with the variation in interest across titles or within titles over time, implicitly assuming that a title's search volume is indicative of demand shocks but is not affected by the timing of piracy protection, at least before protection begins. I find no correlation across titles in overall search volume and the timing of protection. I also find no correlation across time in search volume and the time of protection.

To test whether the decision to have a title protected early is a result of its overall popularity, I plot the title's mean monthly Google Search volume over the observed time frame against the month in which the title moved into protection, with values normalized to a January 2011 search volume of 100 for Arthur C. Clarke's *2001: A Space Odyssey*. Figure 2 shows that, over the observed time period, the mean search volume of titles that were protected earliest (between September and December 2011) was not higher than that of titles that moved into protection later. In addition, the unconditional correlation between the titles' average monthly search volume and the month in which they become protected is positive and close to 0 at .05. This suggests that RosettaBooks did not strategically
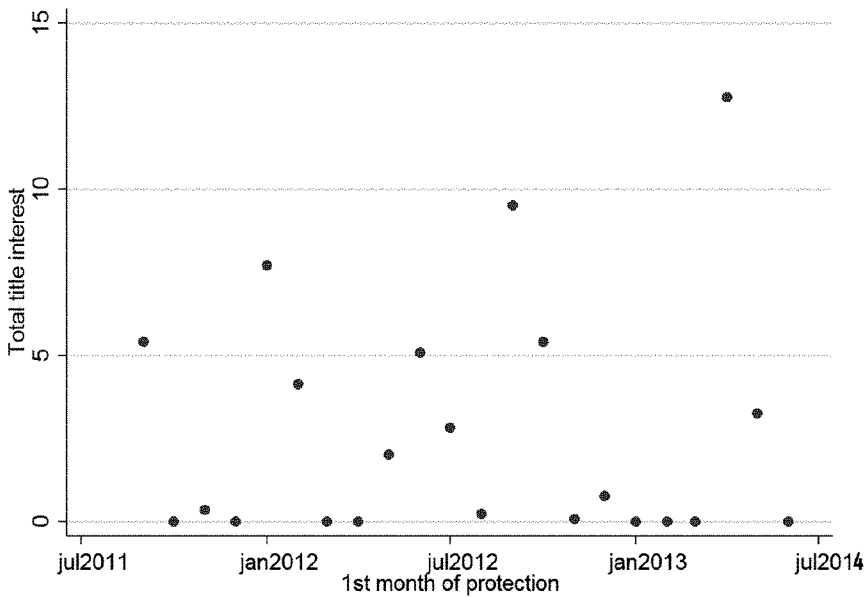
Figure 2. Titles' search volume

move titles for which it expected to extract the highest revenue into protection first.

To test whether the decision to protect a title is driven by sudden spikes in readers' interest in it, I regress each title's search volume in a given month on a quadratic function of the time (in months) before and after it moved into protection. One would expect the coefficient on the linear term to be negative simply because readers' interest likely decreases over time. The coefficient on the quadratic term, on the other hand, provides information about readers' interest level around the start of piracy protection. If readers' interest in a title spikes around the date it moves into protection, one would expect the coefficient on the quadratic term to be negative, but if there is no spike in readers' interest, the coefficient should be close to 0 or even positive.

Table 4 shows that the coefficient of interest is indeed close to 0, or even positive, which suggests that readers' interest is higher in months that are further from protection. This effect is not a result of recent editions being published or title-specific trends, as the regression includes title and month fixed effects and indicators that equal one if an edition of the title was recently published. The pattern holds for title search volume and author search volume and when limiting the range to 6 months before and after the start date (column 1). There is no evidence that readers' interest spikes around the start date of protection.

Table 4
Google Search Volumes for Titles and Authors around the
Start Date of Piracy Protection

	Title			Author
	(1)	(2)	(3)	(4)
Timing	−.0894	−.0326	−.0309	−.118**
	(.0832)	(.0263)	(.0195)	(.0287)
Timing²	.00297	.00169	.00127	.00704**
	(.00296)	(.00164)	(.00124)	(.00260)
Year-month fixed effects	Yes	Yes	No	Yes
Observations	2,234	3,314	3,314	3,314
R^2	.034	.023	.014	.035

Note. Robust standard errors are clustered by title. In column 1 the sample is restricted to the 6 months before and after the start date of protection. All regressions include title fixed effects and controls for recent edition. $N = 126$ titles.
** $p < .01$.

6. The Process of Protection

The above results establish that the informal, targeted type of copyright protection that is commonly used in the book publishing industry can increase legal sales by more than the cost of protection. In what follows, I examine the mechanisms that make this protection effective.

6.1. Measuring the Speed of the Impact of Protection

To analyze the promptness with which protection from piracy takes effect, I employ a regression discontinuity strategy in which I use the time (in months) before and after protection starts (around the cutoff) as the explanatory variable of interest, as in equation (2):

$$\log(q_{it}) = X'_{it}\beta + \left(\sum_{m \in M} \alpha_m \times 1[t - \text{Protection\_start} = m]\right) + \delta_i^k + f(t) + \varepsilon_{it}, \quad (2)$$

where Protection_start is the first month of protection; X_{it} includes the title's age, recent editions, title and author search volume, and the average sales price; δ_i^k includes title format indicators, and $f(t)$ represents a linear time trend.

In Figure 3, month 0 corresponds to the month immediately before protection starts, and month 1 is the first month of protection. The figure shows a statistically significant increase in sales in months 5 and 6 of protection, which indicates that it takes quite some time to find and deactivate enough infringing content. This also suggests that the age and popularity of a title plays an important role when determining the effectiveness of protection. Piracy is likely to reemerge more slowly for older titles, so the antipiracy efforts have a chance to catch up with pirates. Piracy protection is most effective for titles whose demand has leveled off, as new illegal content is less likely to emerge.

Figure 3 provides additional evidence of the exogeneity of piracy protection.
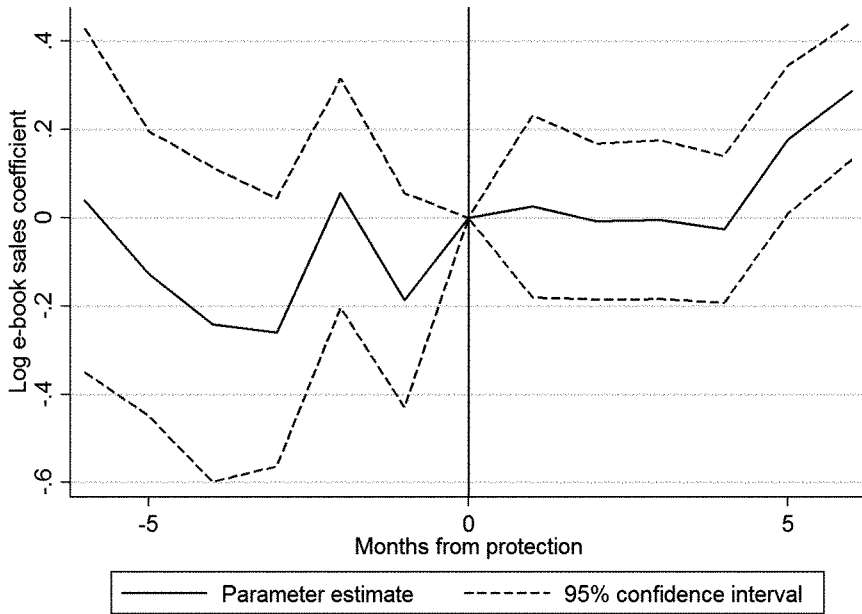
The Journal of LAW & ECONOMICS

Figure 3. Protection and e-book sales

If piracy protection is endogenous, demand for the title will likely be unusually large in the time periods leading up to piracy protection. This does not appear to be the case for the e-books (or other formats) in my data set, as the coefficient on the log of unit sales leading up to protection is close to 0 and does not vary significantly before protection starts.[22]

6.2. Separating the Effects of Delistings and Takedowns

The above estimates show the combined effect of two actions: delisting links from search engines and taking down infringing content on piracy sites. To examine which of these two actions makes piracy protection effective, I add another explanatory variable to equation (1): the direction of the change in the number of active sites for title i compared with the previous time period. A decrease in

---

[22] To alleviate remaining concerns, I repeat this analysis for titles that were originally published more than 20 years ago—titles for which a sudden spike in demand is less likely—and obtain almost identical results. I also estimate the regression with year-month indicators instead of a linear time trend, and the results remain unchanged. In addition, I add linear pre- and postprotection trends, and I add title-specific linear time trends in other regressions. The effect of protection remains robust in direction, magnitude, and statistical significance. Finally, in Section A2 I report the results of a placebo test in which I artificially set the start date of protection several months before the actual start date and show that the effect of protection disappears.

Private Copyright Protection   429

Table 5
Effect of Delistings and Takedown Notices

		Paperback			
	Hardcover	Trade	Mass Market	Audio	E-Book
Delistings	.121	−.137	.156	.223	.147*
	(.352)	(.182)	(.155)	(.394)	(.0613)
Successful takedowns	.0501	−.0462	.127	.0536	.0139
	(.124)	(.0863)	(.110)	(.179)	(.0738)
Log price					−1.482**
					(.155)
Title search volume	No	No	No	No	Yes
Author search volume	No	No	No	No	Yes
Time fixed effects	Weekly	Weekly	Weekly	Weekly	Monthly
Observations	9,350	13,506	8,714	6,630	3,049
R^2	.035	.046	.047	.060	.480
Titles	60	88	59	43	121

Note. Robust standard errors are clustered by title. Results for the dependent variables are in logs of unit sales. All regressions include title fixed effects and controls for recent edition, title age, and active site. Controls for search volume are included for e-books because search volume information is available only at the monthly level for many titles.
* $p < .05$.
** $p < .01$.

the number of active sites indicates that takedown notices were successful in that time period, while an increase in active sites would indicate that more pirated content emerged than was removed. In that case, the only successful component of protection is the delisting of links from search engines. I further control for the overall supply of piracy sites by including the number of detected sites with infringing content that has not yet been taken down for each title in addition to the controls used in Section 5.2.

The relative effect of takedown notices and delistings can be inferred from the relative size and significance of the two coefficients of interest. When a title is under contract to be protected but more content emerges than is taken down, any effect must be due to the delisting of links. If more infringing content is successfully removed than emerges, the estimated effect is a combination of delistings and takedowns. Table 5 reports the effects of each action. It suggests that delistings have a stronger effect on e-book sales than takedowns do. The coefficient on delistings remains almost unchanged compared with the results in Table 3, increasing e-book sales by 15.8 percent. On the other hand, a successful attempt to take down infringing content (a decrease in the number of sites for a title) leads to a small and imprecisely estimated additional increase (1.4 percent) in e-book sales. These results are not surprising, as takedown notices are successful for only 8.7 percent of the title-month combinations under protection.

The relative success of delistings suggests that casual pirates are more easily deterred. A copyright protection strategy that targets casual pirates may be more effective than one that targets more serious pirates. The difference between the

The Journal of LAW & ECONOMICS

Table 6
Effect of Priority Levels on Sales

	Hardcover	Paperback		Audio	E-Book
		Trade	Mass Market		
Low	−.381	−1.010*	−.937	−.0247	.0748
	(.242)	(.397)	(.932)	(.506)	(.0837)
Medium	.207	−.328	−.915	.929+	.0165
	(.632)	(.357)	(.741)	(.483)	(.0824)
High	.00515	−.109	.544**	−.188	.407**
	(.251)	(.159)	(.194)	(.377)	(.150)
Log price					−1.476**
					(.155)
Title search volume	No	No	No	No	Yes
Author search volume	No	No	No	No	Yes
Time fixed effects	Weekly	Weekly	Weekly	Weekly	Monthly
Observations	9,350	13,506	8,714	6,630	3,049
R^2	.035	.051	.062	.064	.483
Titles	60	88	59	43	121

Note. Robust standard errors are clustered by title. Results for the dependent variables are in logs of unit sales. Low-priority titles are searched for once a month, medium-priority titles are searched for once a week, and high-priority titles are searched for daily. All regressions include title fixed effects and controls for recent edition, title age, and active site. Controls for search volume are included for e-books because search volume information is available only at the monthly level for many titles.

$^+ p < .1.$
$^* p < .05.$
$^{**} p < .01.$

coefficients is insignificant at the 10 percent level, however, and the interaction of both actions may be critical for the success of piracy protection. Again, neither type of protection significantly affects physical book sales.[23]

6.3. Intensity of Piracy Protection

Digimarc protects titles with three levels of priority as guided by the observed level of piracy. Websites are crawled approximately once a month for titles that are assigned a low priority, while medium-priority protection corresponds to crawling about once per week, and high priority means the title is searched for every day. One might assume that high-priority titles are about seven times as costly to protect as medium-priority titles. Do those differences in costs outweigh differences in effectiveness?

Table 6 shows that daily crawling has a much larger positive effect than more sporadic crawling. When a title is protected with high priority, e-book sales increase significantly, with a point estimate of 50.2 percent. The effect for medium-priority protection is an order of magnitude smaller. The trend holds true for mass-market paperbacks—a format that can be seen as a close substitute for on-

---

[23] The findings are robust to different definitions of successful takedowns. I show in Section A3 that the results remain almost identical when I use a measure of site traffic instead of the number of sites with infringing content to determine whether takedown notices are successful.

line piracy—and to some extend for the other formats as well. High-intensity protection is at least the most effective, and it might be the most (cost-)efficient as well.

Of course, the decision to assign a certain priority level to a book in a given time period introduces another type of endogeneity. The more a title's copyright is infringed, the more regularly the antipiracy company will search for it, and the more effort its protection requires. In the analysis, I control for the level of piracy (the number of active sites) and for the overall level of readers' interest in the title and author (Google search volume) in addition to the controls for recent editions and the title's age. The fact that the difference is large even after controlling for the various measures of piracy and overall interest suggests that high-frequency crawling indeed is more effective than low-frequency crawling. Titles that are protected more intensely benefit more from protection than those that are searched for less frequently.

6.4. Different Types of Titles

The above results show the overall effects of piracy protection on unit sales and the mechanisms behind such effects, but book piracy can impact different titles in different ways. The effectiveness of piracy protection depends on the title's popularity and the readers it attracts. I explore these differences here. I first analyze the ways in which the effect of piracy protection depends on the title's popularity. I then explore the role of genre.

Piracy affects legal sales through two counteracting channels. The illegal versions can steer consumers away from legal options (displacement effect), but the consumer may also be more likely to hear about a product if free versions are more readily available, which potentially increases legal sales (promotional effect). The relative extent of these effects depends on the past success of a title and its (perceived) quality. If a title has been well known for many years, it does not rely on an additional promotional effect of pirated versions. For well-known and successful titles, book piracy mostly displaces sales of legal editions. On the other hand, titles that are not very well known rely more heavily on word-of-mouth advertising. For those titles, a free pirated version of the book could spur demand for the title by making it accessible to more people. My results support this idea in part, although the displacement effect outweighs the promotional effect for all types of works.

I use the number of Goodreads reviews to proxy for how well known a title is. For instance, Stephen Covey's *The Seven Habits of Highly Effective People* has 162,647 reviews on Goodreads, while the median e-book title in my data set has 483 reviews.[24] I use this variation across titles to separately identify the two effects by interacting the protection term with an indicator that equals one if the work

---

[24] These are the numbers of reviews as of August 2, 2014.

Table 7
Effect of Piracy on Displacement and Promotion

	Hardcover	Paperback		Audio	E-Book
		Trade	Mass Market		
Popular	.227	−.146	.440+	.421	.185+
	(.426)	(.217)	(.239)	(.436)	(.106)
Obscure × Good	−.182	−.149	.861**	−.245	.103
	(.378)	(.356)	(.151)	(.359)	(.0856)
Obscure × Bad	−.167	−.114	−1.994*	−.912	.131
	(.408)	(.926)	(.991)	(1.036)	(.0818)
Log price					−1.480**
					(.153)
Title search volume	No	No	No	No	Yes
Author search volume	No	No	No	No	Yes
Time fixed effects	Weekly	Weekly	Weekly	Weekly	Monthly
Observations	9,350	13,506	8,714	6,630	3,049
R^2	.035	.046	.066	.062	.480
Titles	60	88	59	43	121

Note. Robust standard errors are clustered by title. Results for the dependent variables are in logs of unit sales. All regressions include title fixed effects and controls for recent edition, title age, and active site. Controls for search volume are included for e-books because search volume information is available only at the monthly level for many titles.

+ $p < .1$.
* $p < .05$.
** $p < .01$.

is well known in equation (1). I treat a title as well known if it has more than 500 reviews and as obscure if it has fewer reviews.[25]

The size and direction of this promotional effect depend on the title's quality. If the title is considered to be good (its readers like it), the promotional effect is expected to be positive and rather strong. If, on the other hand, the title is considered to be bad (its readers do not like it), a reader may discourage others from buying the book. This is especially likely for more obscure titles. I interact a measure of quality with the protection variables for the obscure titles in a triple-differences analysis. I treat a title as good if its Goodreads rating is above 3.8—the mean of the titles' ratings across all formats.

Table 7 summarizes the effect of piracy protection on well-known titles and on those that are more obscure. As expected, the effect on e-book sales of popular titles is larger than that on more obscure titles, although the difference is not significant at the 10 percent level. Sales of popular titles increase by 20.3 percent, while sales of more obscure titles increase by just over 10 percent, although this effect is estimated less precisely. This difference supports the notion that there may be a promotional effect for obscure titles, while sales of well-known titles mostly face a displacement effect of piracy. The work's quality, in contrast, does not significantly change the effect of piracy protection, perhaps because informa-

[25] Using Goodreads reviews rather than sales numbers as a measure of popularity helps avoid one source of endogeneity, as sales are the outcome of interest.

tion about quality is readily available through the Internet (for instance, through Goodreads). The results are robust to different cutoff points for popularity and quality.

Finally, while a title's popularity is correlated with the priority with which it is protected, these two measures are not perfectly correlated: .25 percent of the low-priority searches are for well-known titles (according to my measure), and 11 percent of the high-priority searches are for obscure titles. This suggests that the intensities in Section 6.3 capture more than just systematic differences due to the titles' popularity. Moreover, comparing the differences in the effectiveness of protection in Section 6.3 and here, one can see that high-frequency protection increases sales beyond what would be expected simply from being a more popular title. Given the large differences, it seems plausible that daily protection is the most efficient.

The effect of piracy protection further depends on the types of consumers that the work attracts. Some genres benefit more than others from word-of-mouth advertising. In the extreme case, the demand for textbooks likely remains constant over time (as long as the number of students remains unchanged), so piracy mostly displaces book sales without creating much additional demand. On the other extreme, the appeal of a suspense novel by an unknown author is very unclear ex ante.

I observe data for several fiction and nonfiction books.[26] If demand for nonfiction is more certain than that for fiction, one would expect the effect of piracy protection to be larger for nonfiction works. Table 8 shows the effect of piracy protection on both genres. The estimation now includes interaction terms of piracy protection and indicator variables for the two types of works. Editions of nonfiction titles benefit more from piracy protection than those of fiction titles for all formats, although the difference is statistically insignificant at the 10 percent level throughout.

7. Conclusion and Policy Implications

Digitization has significantly lowered the cost of production and distribution of cultural goods over the past decade. This has resulted in large shifts in market structure and competitive behavior, which has led to an increase in the variety of products that can be consumed. Some of these new products infringe the copyrights of existing works. Economists have shown that file sharing and online piracy have had large impacts on media industries. As a result, making the protection of a copyright in such industries effective and efficient is a focus among academics and policy makers alike.

While current issues surrounding copyright pertain mostly to software and media goods, questions about piracy and copyright enforcement will likely gain importance in the future as more products are digitized, for example, via 3-D

---

[26] The list of RosettaBooks' titles includes more fiction than nonfiction works: 85 percent of the e-books and 74 percent of the physical editions in the analysis are fiction.

Table 8
Effect of Piracy Protection by Genre

	Hardcover	Paperback		Audio	E-Book
		Trade	Mass Market		
Fiction	−.192	−.236	.129	.214	.138*
	(.275)	(.244)	(.148)	(.443)	(.0663)
Nonfiction	.863	.0537	.524+	.275	.215*
	(.748)	(.222)	(.308)	(.860)	(.0104)
Log price					−1.483**
					(.155)
Title search volume	No	No	No	No	Yes
Author search volume	No	No	No	No	Yes
Time fixed effects	Weekly	Weekly	Weekly	Weekly	Monthly
Observations	9,350	13,506	8,714	6,630	3,049
R^2	.038	.047	.048	.060	.480
Titles	60	88	59	43	121

Note. Robust standard errors are clustered by title. Results for the dependent variables are in logs of unit sales. All regressions include title fixed effects and controls for recent edition, title age, and active site. Controls for search volume are included for e-books because search volume information is available only at the monthly level for many titles.
  + $p < .1$.
  * $p < .05$.
  ** $p < .01$.

printing. One of the biggest issues is who should carry the burden of protecting copyrighted works—government, the firms selling the works, or other firms (for example, private firms, Internet service providers, or search engines).

Thus far, most efforts have been legislative or at least formal and broad in nature, with mixed reception and success. Many governmental (legislative) attempts have faced significant opposition among the public, with several bills being shelved before their introduction (for example, SOPA and PIPA in the United States) or partially revoked (for example, HADOPI in France). Although legal efforts have been less controversial, they can be quite costly, as many efforts were intended to protect entire industries (for example, the shutdown of Megaupload), and most involved expensive lawsuits.

The book publishing industry demonstrates a rather different approach—one in which private companies provide surveillance of individual works. This type of protection is less likely to be controversial, and costly court proceedings can largely be avoided. Whether piracy protection is effective in this setting may indicate whether private action can prevent copyright infringement. Even more, if we know what types of works are most likely to benefit from copyright protection, policy makers and enforcers can make protection more cost-efficient. This paper offers insights into the cost-effectiveness of protection on three dimensions: the types of pirates offering and consuming these works, the effort levels used to protect works from piracy, and the types of works that are protected.

First, the book publishing industry may be particularly well suited for piracy

protection. The small file sizes of digital books compared with other media prod-
ucts make offering and consuming pirated content particularly easy. As a result,
one might expect more casual pirates on both the demand and the supply sides
of piracy. Casual suppliers are more likely to be intimidated into removing the
content, as can be seen in the high success rate of takedown notices for content
on cyberlockers and general websites compared with peer-to-peer networks. On
the demand side, much of the effectiveness of piracy protection can be attributed
to the removal of links to infringing content on search engines—presumably a
search channel most often used by casual pirates. Since a request to delist a link
from search engines can be sent relatively easily, a more efficient copyright pro-
tection strategy could target these channels rather than attempt to have the con-
tent removed, especially since such attempts can become quite costly if they lead
to legal escalation.

Second, protection has the largest effect when it is intense and infringing con-
tent is searched for, delisted, and removed frequently. The differences in effec-
tiveness may outweigh differences in costs, although a direct comparison is be-
yond the scope of this study.

Finally, for well-known works and those by popular authors, online piracy
poses a threat mainly to physical book sales, while authors who are less estab-
lished may see a promotional effect in addition to the displacement effect (see
Peitz and Waelbroeck 2006b). My results support this idea, although the dis-
placement effect of piracy outweighs the promotional effect regardless of the type
of work.[27]

I analyze a particular set of titles: those originally published several years ago
and not necessarily the current top sellers. The dynamics that affect demand are
different for recently published titles than for titles that have been available for
a long time. The time it takes for protection to have an effect is especially long
when considering the sales patterns of recently published works, as titles typi-
cally are most successful right after publication. Piracy protection may be most
effective for works whose sales have leveled off and where new piracy is less likely
to emerge quickly. Future research should therefore examine whether copyright
protection can also be effective for more recent works.

Appendix

Additional Robustness Checks

My main results are robust to small changes in the definition of the success of
protection. They are also robust to specifications that include interactions of time
fixed effects with dummy variables for types of titles. In what follows, I present in
more detail results that have different interpretations of protection and its suc-

---

[27] Nevertheless, the promotional effect of free online versions is assumed by emerging authors
who turn to offering free excerpts of their titles on their own websites. This strategy allows authors
to control the level of sampling and promotion, whereas they have little to no control over online
piracy.

Table A1
Effect of Sporadic Piracy Protection

	Hardcover	Paperback		Audio	E-Book
		Trade	Mass Market		
Protection	.0476	.00358	.236*	−.0122	.132**
	(.167)	(.0921)	(.0874)	(.178)	(.0445)
Log price					−1.483**
					(.156)
Title search volume	No	No	No	No	Yes
Author search volume	No	No	No	No	Yes
Time fixed effects	Weekly	Weekly	Weekly	Weekly	Monthly
Observations	9,350	13,506	8,714	6,630	3,049
R^2	.034	.046	.048	.059	.480
Titles	60	88	59	43	121

Note. Robust standard errors are clustered by title. Results for the dependent variables are in logs of unit sales. Protection is in effect if content on at least one infringing site is removed. All regressions include title fixed effects and controls for recent edition, title age, and active site. Controls for search volume are included for e-books because search volume information is available only at the monthly level for many titles.
  * $p < .05$.
  ** $p < .01$.

cess and the results of a placebo test on the timing of protection. The result for the effectiveness of protection is robust to these alternative specifications.

A1. Robustness Check: Protection When Infringing Content Is Removed

My main analysis assumes that a title is constantly protected. That is, Digimarc constantly searches the Internet for infringing content and constantly removes the links to it. However, Digimarc searches the Internet for some of the titles more sporadically (once a week or once a month). It is also possible that Digimarc does not find some of the links to infringing content. A title is protected only when the company actively and successfully tries to limit piracy.

Table A1 shows the estimation results when protection is in effect during the time periods when infringing content is taken down. This is a stricter definition of piracy protection, so larger effects may be expected.

While the effect remains insignificant for most physical formats, it becomes significant (at 5 percent) for mass-market paperbacks. This format constitutes the cheap, low-quality paperback versions that can be found at airport bookstores. These are closer substitutes for free pirated versions than, for example, hardcover editions, and the significant positive effect is not surprising. Interestingly, the effect on e-books is smaller than it is under the assumption of constant protection. The effect remains largely significant, however, which indicates again that the estimates in the main analysis are consistent. Since titles move in and out of protection, there is less concern about path dependency of the treatment variable.

Table A2
Placebo Test: E-Book Sales with Artificial Start Dates for Piracy Protection

	(1)	(2)	(3)	(4)	(5)
Protection	.0893	.146	.0420	−.0178	.0348
	(.100)	(.0970)	(.0862)	(.0873)	(.102)
Log price	−1.185**	−1.178**	−1.185**	−1.339**	−1.709**
	(.165)	(.163)	(.166)	(.256)	(.283)
Months early	2	3	4	5	6
Observations	1,233	1,227	1,215	1,158	1,111
R^2	.442	.444	.444	.447	.472
Titles	102	100	97	83	75

Note. Robust standard errors are clustered by title. Results for the dependent variables are in logs of unit sales. All regressions include controls for recent edition, title age, active site, and title and author search volumes.

** $p < .01$.

A2. Robustness Check: Placebo Test

Despite the controls in the main estimation, there may be concern that the results pick up additional effects that are not accounted for and remain unobservable. Figure 3 suggests that unit sales right before protection begins are not significantly higher than those in the preceding months, although the coefficients are estimated quite imprecisely.

Here I address these concerns more directly by testing whether books that are about to move into protection already see an increasing sales trend compared with books that move into protection later. I exclude all data points after the title goes into protection and set the placebo protection date 2–6 months before the actual protection date. The analysis then is the same as in equation (1). Table A2 shows that while the coefficient on placebo protection is positive for most artificial start dates, it is much less so than in the main specification and never statistically significant. Even if this estimation were to suggest an upward trend in sales prior to the start date of protection, this trend does not explain the large, statistically significant effect that is estimated in the main specification. Protection remains cost-efficient when comparing this net effect to the costs of protection.

A3. Robustness Check: Site Traffic as a Measure of Success

Section 6.2 uses the change in the number of available infringing sites to identify the effects of delistings and takedown notices. It is possible, though, that those sites with infringing content that is successfully removed do not receive a lot of traffic anyway. In that case, takedown notices are not as successful as they appear.

I proxy for traffic to the infringing content by using site-ranking information from Alexa Internet, Inc., as retrieved on September 25, 2014. This company provides traffic data, global rankings, and other information about millions of websites (the lowest observed ranking in my data set is 18,943,656). Note that
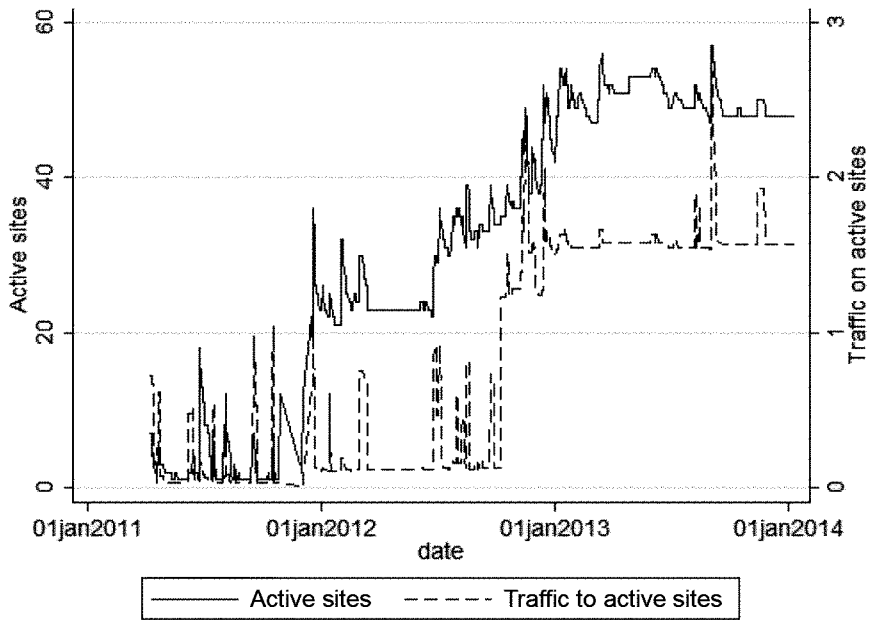
Figure A1. Active piracy sites and estimated traffic for *I Am Legend*

Table A3

Effect of Delistings and Takedowns: Traffic to Infringing Content

| | | Paperback | | | |
	Hardcover	Trade	Mass Market	Audio	E-Book
Under contract	.0951	−.137	.182	.249	.135*
	(.350)	(.183)	(.145)	(.395)	(.0596)
Successful takedowns	.0587	.00624	.00966	.00321	.0231
	(.126)	(.0762)	(.0918)	(.161)	(.0722)
Traffic to active sites	.131	.0899	.337	−.433	.119
	(.421)	(.168)	(.207)	(.444)	(.188)
Log price					−1.484**
					(.155)
Time fixed effects	Weekly	Weekly	Weekly	Weekly	Monthly
Observations	9,350	13,506	8,714	6,630	3,049
R^2	.033	.046	.048	.060	.480
Titles	60	88	59	43	121

Note. Robust standard errors are clustered by title. Results for the dependent variables are in logs of unit sales. Takedowns are considered successful if the imputed site traffic to the infringing content is lower at the end of the weekly or monthly time period than at the end of the previous time period. All regressions include title fixed effects and controls for recent edition and title age. Controls for search volume are included for e-books because search volume information is available only at the monthly level for many titles.

* $p < .05$.

** $p < .01$.

the sites reported on Alexa.com are the host sites to the infringing links, which means that I do not observe traffic to the infringing content itself.

I use the inverse of the host site's ranking, assuming that traffic on the Internet follows a Pareto distribution and that traffic to the infringing content is proportional to traffic to the main site. Figure A1 shows that traffic to infringing sites is not perfectly correlated with the number of active infringing sites. Here I estimate the two effects, treating takedown notices as successful if the estimated traffic to infringing content for a title, rather than the number of infringing sites, decreases in a given time period.

Table A3 shows that the relative effects of delisting a site and of a successful attempt to remove infringing content are similar to those found when using the number of active infringing sites (Table 5). The effect of delistings again appears to be much stronger, although the difference remains statistically insignificant at the 10 percent level.

References

Aguiar, Luis, Jörg Claussen, and Christian Peukert. 2015. Online Copyright Enforcement, Consumer Behavior, and Market Structure. Working Paper No. 2015/01. European Commission, Institute for Prospective Technological Studies, Seville.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics* 119:249–75.

Bhattacharjee, Supid, Ram D. Gopal, Kaveepan Lertwachara, and James R. Marsden. 2006. Impact of Legal Threats on Online Music Sharing Activity: An Analysis of Music Industry Legal Actions. *Journal of Law and Economics* 49:91–114.

Danaher, Brett, and Michael D. Smith. 2014. Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales. *International Journal of Industrial Organization* 33:1–8.

Danaher, Brett, Michael D. Smith, and Rahul Telang. 2015. The Effect of Piracy Website Blocking on Consumer Behavior. Working paper. Carnegie Mellon University, Department of Economics, Pittsburgh.

Danaher, Brett, Michael D. Smith, Rahul Telang, and Siwen Chen. 2014. The Effect of Graduated Response Anti-piracy Laws on Music Sales: Evidence from an Event Study in France. *Journal of Industrial Economics* 62:541–53.

Deahl, Rachel. 2011. Rosetta Books Comes of Age. *Publishers Weekly,* July 11. http://www .publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/47933 -rosetta-books-comes-of-age.html.

Donald, Stephen G., and Kevin Lang. 2007. Inference with Difference-in-Differences and Other Panel Data. *Review of Economics and Statistics* 89:221–33.

Hardy, Wojciech, Michał Krawczyk, and Joanna Tyrowicz. 2014. Internet Piracy and Book Sales: A Field Experiment. Working Paper No. 23. University of Warsaw, Faculty of Economics, Warsaw.

Liebowitz, Stan J. 2008. Research Note: Testing File Sharing's Impact on Music Album Sales in Cities. *Management Science* 54:852–59.

Luo, Hong, and Julie Holland Mortimer. 2015. Copyright Enforcement: Evidence from

Two Field Experiments. Working Paper No. 16-024. Harvard Business School, Boston.

Oberholzer-Gee, Felix, and Koleman Strumpf. 2007. The Effect of File Sharing on Record Sales: An Empirical Analysis. *Journal of Political Economy* 115:1–42.

Oestreicher-Singer, Gal, and Arun Sundararajan. 2010. Are Digital Rights Valuable? Theory and Evidence from Ebook Pricing. Working Paper No. 06-01. New York University, Center for Digital Economy Research, New York.

Peitz, Martin, and Patrick Waelbroeck. 2006a. Piracy of Digital Products: A Critical Review of the Theoretical Literature. *Information Economics and Policy* 18:449–76.

———. 2006b. Why the Music Industry May Gain from Free Downloading—the Role of Sampling. *International Journal of Industrial Organization* 24:907–13.

Petrin, Amil. 2002. Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy* 110:705–29.

Peukert, Christian, Jörg Claussen, and Tobias Kretschmer. 2013. Piracy and Movie Revenues: Evidence from Megaupload: A Tale of the Long Tail? Working paper. Ludwig Maximilian University of Munich, Faculty of Economics, Business Administration and Information Technology, Munich.

Rob, Rafael, and Joel Waldfogel. 2006. Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students. *Journal of Law and Economics* 49:29–62.

Sivan, Liron, Michael D. Smith, and Rahul Telang. 2014. Do Search Engines Influence Media Piracy? Evidence from a Randomized Field Study. Working paper. Carnegie Mellon University, Heinz School of Public Policy and Management, Pittsburgh.

Smith, Michael D., and Rahul Telang. 2012. Assessing the Academic Literature regarding the Impact of Media Piracy on Sales. Working paper. Carnegie Mellon University, Heinz School of Public Policy and Management, Pittsburgh.

Sterling, Greg. 2014. Pew: 50 Percent in US Now Own Tablet or E-Reader. *Marketing Land,* January 16.

Waldfogel, Joel. 2010. Music File Sharing and Sales Displacement in the iTunes Era. *Information Economics and Policy* 22:306–14.

Waldfogel, Joel, and Imke Reimers. 2015. Storming the Gatekeepers: Digital Disintermediation in the Market for Books. *Information Economics and Policy* 31:47–58.

Zentner, Alejandro. 2006. Measuring the Effect of File Sharing on Music Purchases. *Journal of Law and Economics* 49:63–90.

Zhang, Laurina. 2014. Intellectual Property Strategy and the Long Tail: Evidence from the Recorded Music Industry. Working paper. Western University, Ivey Business School, London, ON.